The Claims

1. (Currently amended) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

store data indicative of a data transmission rate for each of previous server operations;

generate, by averaging the stored data transmission rates, a bandwidth value indicating an average bandwidth used by a server to perform the previous server operations;

receive a request for the server to perform an additional server operation;

compare the bandwidth value to a threshold; and

restrict the request in a first manner if the bandwidth value exceeds the threshold.

2. (Canceled).

3. (Currently amended) One or more computer-readable media ~~as recited in claim 1, wherein the plurality of instructions cause the one or more processors to~~ having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

generate a bandwidth value indicating an average bandwidth used by a server to perform previous server operations;

receive a request for the server to perform an additional server operation;

compare the bandwidth value to a threshold; and

restrict the request in ~~the~~ a first manner if both the bandwidth value exceeds the threshold and the additional server operation is of a first type.

4. (Original) One or more computer-readable media as recited in claim 3, wherein the first type comprises a read operation.

5. (Original) One or more computer-readable media as recited in claim 3, wherein the first type comprises a write operation.

6. (Original) One or more computer-readable media as recited in claim 3, wherein the first type comprises a transmit operation.

7. (Original) One or more computer-readable media as recited in claim 1, wherein the previous server operations and the additional server operation include one or more of read operations, write operations, and transmit operations.

8. (Currently amended) One or more computer-readable media ~~as recited in claim 1, wherein the plurality of instructions further cause the one or more processors to~~ having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

generate a bandwidth value indicating an average bandwidth used by a server to perform previous server operations;

receive a request for the server to perform an additional server operation;

compare the bandwidth value to a threshold;

restrict the request in a first manner if the bandwidth value exceeds the threshold;

compare the bandwidth value to another threshold; and

restrict the request in a second manner if the bandwidth value exceeds the another threshold.

9. (Currently amended) One or more computer-readable media as recited in claim 8, wherein:

the threshold is less than the another threshold;

restricting the request in the first manner comprises delaying the request; and

restricting the request in the second manner comprises blocking the request.

10. (Original) One or more computer-readable media as recited in claim 1, wherein the restricting the request in the first manner comprises delaying the request.

11. (Original) One or more computer-readable media as recited in claim 1, wherein the restricting the request in the first manner comprises blocking the request.

12. (Original) One or more computer-readable media as recited in claim 1, wherein the restricting the request in the first manner comprises preventing the request from being presented to the server.

13. (Original) One or more computer-readable media as recited in claim 1, wherein the restricting the request in the first manner comprises preventing the request from entering a request queue of the server.

14. (Currently amended) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to; ~~as recited in claim 1, wherein the plurality of instructions include instructions that cause the one or more processors to generate the bandwidth value~~

generate a bandwidth value indicating an average bandwidth used by a server to perform previous server operations by:

generating a value for each of the previous server operations by,

identifying a time interval duration between a start time of the previous server operation and an ~~the~~ end time of the previous server operation,

identifying a number of bytes transferred for the previous server operation, and

dividing the number of bytes by the time interval duration; and

dividing a sum of the values of the previous server operations by the number of previous server operations;

receive a request for the server to perform an additional server operation;

compare the bandwidth value to a threshold; and

restrict the request in a first manner if the bandwidth value exceeds the threshold.

15. (Currently amended) A method comprising:

receiving a request to perform a server operation; and

restricting performance of the request based at least in part on an average bandwidth used in performing previous server operations, wherein the restricting comprises:

comparing the average bandwidth to at least one of a first threshold and a second threshold;

restricting the request in a first manner if the average bandwidth exceeds the first threshold but does not exceed the second threshold; and

restricting the request in a second manner if the average bandwidth exceeds the second threshold.

16. (Original) A method as recited in claim 15, further comprising:

storing data indicative of a data transmission rate for each of the previous server operations; and

generating the average bandwidth by averaging the stored data transmission rates.

17. (Currently amended) A method as recited in claim 15, wherein the restricting further comprises restricting the request in the first manner if both the average bandwidth ~~value~~ exceeds [[a]] the first threshold and the server operation is of a first type.

18. (Original) A method as recited in claim 17, wherein the first type comprises a read operation.

19. (Original) A method as recited in claim 17, wherein the first type comprises a write operation.

20. (Original) A method as recited in claim 17, wherein the first type comprises a transmit operation.

21. (Canceled).

22. (Currently amended) A method as recited in claim 15~~21~~, wherein restricting the request in [[a]] the first manner comprises delaying the request and wherein restricting the request in [[a]] the second manner comprises blocking the request.

23. (Canceled).

24. (Currently amended) A host system comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to determine whether performance of the request by the one of the at least one network server is to be restricted based at least in part on an average bandwidth used by the one of the at least one network server in performing previous server operations, wherein the bandwidth throttling system is further to:

store data indicative of a data transmission rate for each of the previous server operations; and

generate the average bandwidth by averaging the stored data transmission rates.

25. (Canceled).

26. (Original) A host system as recited in claim 24, further comprising an ancillary function driver, coupled to the asynchronous thread queue, to couple the host system to a network.

27. (Currently amended) A host system ~~as recited in claim 24,~~ comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to determine whether performance of the request by the one of the at least one network server is to be restricted based at least in part on an average bandwidth used by the one of the at least one network server in performing previous operations;

wherein the at least one network server comprises a plurality of network servers, and wherein the bandwidth throttling system determines whether performance of a request by a particular one of the plurality of network servers is to be restricted based on the average bandwidth used by that particular network server in performing previous operations and independent of the average bandwidth used by other network servers of the plurality of network servers in performing previous operations.

28. (Currently amended) A host system ~~as recited in claim 24,~~ comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to determine whether performance of the request by the one of the at least one network server is to be restricted based at least in part on an average bandwidth

used by the one of the at least one network server in performing previous operations;

wherein the bandwidth throttling system comprises:

a measurement subsystem to compute the average bandwidth used by the one of the at least one network server in performing previous operations; and

a control subsystem, coupled to the measurement subsystem, to make the determination and communicate how the asynchronous thread queue is to restrict performance of the request.

29. (Currently amended) A method comprising:

storing data indicative of a data transmission rate for each of previous server operations;

generating, by averaging the stored data transmission rates, a value indicating a bandwidth used by a server to perform the previous server operations;

receiving a request for the server to perform an additional server operation; and

restricting the request in a first manner if the value exceeds a first threshold but does not exceed a second threshold, and restricting the request in a second manner if the value exceeds the second threshold.

30. (Canceled).

31. (Currently amended) A method ~~as recited in claim 29, further~~ comprising:

generating a value indicating a bandwidth used by a server to perform previous server operations;

receiving a request for the server to perform an additional server operation;

restricting the request in a first manner if the value exceeds a first threshold but does not exceed a second threshold, and restricting the request in a second manner if the value exceeds the second threshold;

generating another value indicating bandwidth used by another server to perform other previous server operations;

receiving another request for the another server to perform another additional server operation; and

restricting the other request in the first manner if the value exceeds a third threshold but does not exceed a fourth threshold, and restricting the other request in the second manner if the value exceeds the fourth threshold.

32. (Currently amended) A method as recited in claim 3129, wherein the first threshold is different than the third threshold, and wherein the second threshold is different than the fourth threshold.

33. (Original) A method as recited in claim 29, wherein the first manner comprises delaying the request.

34. (Original) A method as recited in claim 29, wherein the second manner comprises blocking the request.

35. (Original) A method as recited in claim 29, wherein the restricting the request in the first manner comprises preventing the request from being presented to the server.

36. (Original) A method as recited in claim 29, wherein the restricting the request in the first manner comprises preventing the request from entering a request queue of the server.

37. (Original) A method as recited in claim 29, wherein the value indicates an average bandwidth used by the server.

38. (Original) A method as recited in claim 29, wherein the restricting comprises restricting the request in the first manner or the second manner only if the request is of a first type.

39. (Original) A method as recited in claim 38, wherein the first type comprises a read operation.

40. (Original) A method as recited in claim 38, wherein the first type comprises a write operation.

41. (Original) A method as recited in claim 38, wherein the first type comprises a transmit operation.

42. (Original) One or more computer-readable memories comprising computer-readable instructions that, when executed by a processor, direct a computer system to perform the method as recited in claim 29.

43. (Currently amended) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

store data indicative of a data transmission rate for each of previous server operations;

generate, by averaging the stored data transmission rates, a value indicating a bandwidth used by a server to perform the previous server operations;

receive a request for the server to perform an additional server operation;

compare the value to at least one of a first threshold and a second threshold; and

delay the request if the value exceeds the first threshold but does not exceed the second threshold, and block the request if the value exceeds the second threshold.

44. (Canceled).

45. (Original) One or more computer-readable media as recited in claim 43, wherein the value indicates an average bandwidth used by the server.

46. (Currently amended) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

generate a value indicating a bandwidth used by a server to perform previous server operations;

receive a request for the server to perform an additional server operation;

compare the value to at least one of a first threshold and a second threshold; and

~~wherein the delaying comprises delaying the request~~ delay the request only if the request is of a first type and if the value exceeds the first threshold but does not exceed the second threshold, and ~~wherein the blocking comprises blocking the request~~ block the request only if the request is of [[a]] the first type and if the value exceeds the second threshold.

47. (Original) One or more computer-readable media as recited in claim 46, wherein the first type comprises a read operation.

48. (Original) One or more computer-readable media as recited in claim 46, wherein the first type comprises a write operation.

49. (Original) One or more computer-readable media as recited in claim 46, wherein the first type comprises a transmit operation.

50. (Canceled).

51. (Currently amended) A host system ~~as recited in claim 50,~~ comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to compare a value indicating a bandwidth used by the one of the at least one network server to perform previous server operations to at least one of a first threshold and a second threshold, to restrict the request in a first manner if the value exceeds the first threshold but does not exceed the second threshold, and to restrict the request in a second manner if the value exceeds the second threshold;

wherein the bandwidth throttling system is further to:

store data indicative of a data transmission rate for each of the previous server operations; and

generate the value by averaging the stored data transmission rates.

52. (Currently amended) A host system ~~as recited in claim 50,~~ comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to compare a value indicating a bandwidth used by the one of the at least one network server to perform previous server operations to at least one of a first threshold and a second threshold, to restrict the request in a first manner if the value exceeds the first threshold but does not exceed the second threshold, and to restrict the request in a second manner if the value exceeds the second threshold;

further comprising a plurality of network servers, and wherein the bandwidth throttling system determines whether to restrict a request for a particular one of the plurality of network servers based on the value indicating bandwidth used by that particular network server in performing the previous server operations and independent of ~~the~~ bandwidth used by other network servers of the plurality of network servers in performing the previous server operations.

53. (Currently amended) A host system ~~as recited in claim 50,~~ comprising:

at least one network server;

an asynchronous thread queue to receive a request, from a client process, to be performed by one of the at least one network server; and

a bandwidth throttling system, coupled to the asynchronous thread queue, to compare a value indicating a bandwidth used by the one of the at least one network server to perform previous server operations to at least one of a first threshold and a second threshold, to restrict the request in a first manner if the

value exceeds the first threshold but does not exceed the second threshold, and to restrict the request in a second manner if the value exceeds the second threshold;

wherein the bandwidth throttling system comprises:

a measurement subsystem to compute, as the value, an average bandwidth used by the network server in performing the previous server operations; and

a control subsystem, coupled to the measurement subsystem, to make the determination and communicate how the asynchronous thread queue is to restrict performance of the request.

54. (Original) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

determine a presently used bandwidth for each of at least one network server by way of a data transmission rate measurement during execution of an operation for each of said at least one network server that includes:

storing data indicative of said data transmission rate measurement of said operation for n last most recently executed operations for each of said at least one network server, wherein n is a positive integer greater than 2, and

generating data indicative of an effective presently used bandwidth for each of said at least one network server wherein said effective presently used bandwidth includes an average of said data transmission rate measurement over said n last most recently executed operations;

effect provision of a plurality of classes of service provided by one of said at least one network server in a first manner, in response to said effective presently used bandwidth for said one of said at least one network server that exceeds a first threshold; and

effect provision of said plurality of classes of service provided by said one of said at least one network server in a second manner that differs from said first manner, in response to said effective presently used bandwidth for said one of said at least one network server that exceeds a second threshold that is greater than said first threshold.

55. (Currently amended) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

store data indicative of a predetermined allocated data transmission bandwidth for each of a plurality of network servers, each of said plurality of network servers providing a plurality of classes of service;

determine an effective bandwidth for each of said plurality of network servers, comprising:

calculating a bandwidth for each operation performed by each of said plurality of network servers,

tabulating a count of bandwidth used by each of said plurality of network servers in each of a plurality of last time intervals, and

averaging said tabulated count of bandwidth to obtain a value indicative of said effective bandwidth;

delay, in response to said effective bandwidth for a one of said network servers substantially corresponding to said predetermined allocated data transmission bandwidth for said one network server, for delaying a first subset of said plurality of classes of service provided by said one network server; and

reject, in response to said effective bandwidth for said one network server exceeding said predetermined allocated data transmission bandwidth for said one network server, requests for said first subset of classes of service provided by said one network server and for delaying service for a second subset of said plurality of classes of service provided by said one network server.

56. (Original) One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

determine a presently used bandwidth for each of a plurality of network servers by way of a data transmission rate measurement taken during execution of an operation for each of said plurality of network servers, said determining including:

storing data indicative of a predetermined allocated data transmission bandwidth for each of said plurality of network servers, and

storing data indicative of a first threshold, wherein said first threshold includes an indication of a differential from said predetermined allocated data transmission bandwidth for each of said plurality of network servers; and

effect provision of a plurality of classes of service provided by a first one of said plurality of network servers in a manner that is individually defined for a plurality of successively greater thresholds in response to said effective presently used bandwidth of said first one of said plurality of network servers exceeding one of said plurality of successively greater thresholds, wherein said manner defined for said first one of said plurality of network servers differs from every other manner defined for other ones of said plurality of network servers.